Guidance to the MAQC Main Study

(August 18, 2005, 3:00 pm CDT)

1. Summary Statement

The MAQC project promises to be the most ambitious and comprehensive study to date on microarray quality control and cross-platform comparison. A consortium of many commercial, government, and academic participants has contributed and will continue to contribute substantial resources, time and expertise to make this study a success. To achieve the intended, long-term benefits of the MAQC main study, guidance is needed before experimental processing of samples begins.

The validity of the study can be seriously compromised if all participants of the study do not operate under the same set of guidelines. We recognize that the number of MAQC participants has been growing rapidly, and recent teleconferences have demonstrated that the level of shared understanding between participants is variable. We propose that a detailed guidance document be drafted to ensure a common basis of understanding for participants of the MAQC study.

2. History of the MAQC Study

The history and background for the MAQC study is best summarized by the website: http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm. Most of the decisions regarding experimental design elements and responsibilities of participants for the main study were decided during the face-to-face meeting in Rockville, MD, May 2-3, 2005 and discussed during subsequent teleconferences. Summaries are available at the MAQC website.

3. Objectives and Design of the MAQC Main Study

3.1 Objectives

1. Measure intra-platform performance. Measure the performance achievable on a microarray platform by laboratories practicing standardized protocols using consistent lots of microarrays and target preparation kits. A set of performance metrics will be defined under the section on Data Analysis Procedures.
2. Measure inter-platform comparability of differential gene expression measurements.
3. Measure relative accuracy as defined by titrated mixtures of the two reference RNA samples.
4. Compare the concordance of expression measurements to other array platforms and alternative technologies (e.g., TaqMan and QuantiGene) based on a pre-defined list of "common genes" that share a common reference sequence to which the probes were designed (see Objective 5).
5. Develop a framework for conducting cross-platform mapping based on probe sequence mapping to the RefSeq database.
6. Generate quality control data about the reference samples before and after target preparation to assist with data analysis and interpretation of the final results.

7. Provide a substantial dataset to serves as a reference to the microarray community for comparing quality, and performance within laboratories.

The following variables were decided to be beyond the scope of the MAQC main study:

- Day-to-day variability
- Operator-to-operator variability
- Variability due to sample quality
- Variability due to the use of different protocols or reagents on the same platform
- Spike-in samples to measure absolute accuracy, sensitivity, specificity, or dynamic range

3.2 Experimental Design

1. Five (5) replicates per site X 4 RNA samples X 3 sites per platform (Table 1), resulting in 20 arrays per site and 60 arrays per platform. (Note: Only the UHRR and brain samples will be processed on the Agilent platform; 6 sites will be testing the Affymetrix platform.)
2. Eight (8) microarray platforms will be tested (Table 2).
3. Sequence-based mapping of probes from different platforms to the RefSeq database will identify a list of genes that have shared local sequences from which the probes were designed. A list of 5,381 RefSeq's have been identified to be targeted by Affymetrix, Agilent, Applied Biosystems (both microarrays and TaqMan assays), Combimatrix, GE Healthcare, Illumina, and NCI_Operon platforms (based on perfect match of probe sequences to the July-25-2005 version of RefSeq database by each platform).
4. Transcripts will be quantified by TaqMan QRT-PCR (1,000 genes), Genospectra (200 genes), and Stratagene (20 genes). The genes are selected from the list of 5,381 RefSeq's to cover different expression levels, fold changes, and some genes showing cross-platform discrepancy in the pilot study. Level of replication and how replication is performed should be defined by Applied Biosystems, Genospectra, and Stratagene in their respective SOP's. Some of the assays (e.g., Stratagene's QRT-PCR and Genospectra's QuantiGene) may be conducted on a subset of "trouble genes" after the microarray data collection is completed.
5. Targets will be prepared together on the same day by the same operator. Platform providers should indicate whether this is within practical capabilities of a single operator. The study design is confounded if targets are prepared by multiple people or on multiple days. If this cannot be achieved, the actual method of target preparation needs to be recorded.

3.3 Four RNA Samples

Ambion's Human Brain Reference RNA (HBRR, Catalog number: TBD), Stratagene's Universal Human Reference RNA (SUHRR, Catalog number: 740000), and two mixtures in the mass ratio of 25%:75% and 75%:25% (brain:SUHRR) will be tested in the MAQC main study (Table 1). The SUHRR will be shipped to Ambion and mixed with Ambion's brain RNA to make the two mixtures. The four RNA samples will be shipped from Ambion to each test site in one single package.

Table 1. Four RNA samples

RNA	Description	Product number
A	Stratagene UHRR (SUHRR)	740000
B	Ambion Human Brain Reference RNA (HBRR)	TBD
C	25% Brain / 75% SUHRR	
D	75% Brain / 25% SUHRR	

3.4 Eight Microarray Platforms

Eight microarray platforms will be tested in the MAQC main study (Table 2).

Table 2. Eight microarray platforms and two alternative technologies

No.	Code	Manufacturer	Product	# of probe(set)s
1	AFX	Affymetrix	U133 Plus2.0	54,675
2	AGL	Agilent	G4112A	43,931
3	ABI	Applied Biosystems	Genome Survey Microarray V2.0	32,878
4	CMB	Combimatrix	CustomArray for MAQC	9,598
5	EPP	Eppendorf	DualChip Microarrays	1,336*
6	GEH	GE Healthcare	300026-6PK	54,167
7	ILM	Illumina	Human-6 Expression BeadChips, 48K v1.0	48,846
8	NCI	NCI_Operon	Operon Human V3	35,328
9 (TaqMan)	TAQ	Applied Biosystems	(TaqMan assays for 7,828 RefSeq's are available)	>=1,000 genes
10 (QuantiGene)	QGN	Genospectra	(~1,272 assays available)	>=200 genes

*Only ~300 genes will be printed and assayed.

3.5 Two Alternative Technologies

Two alternative technologies, *i.e.*, TaqMan QRT-PCR (Applied Biosystems) and QuantiGene (Genospectra) will be tested in the MAQC main study (Table 2).

4. Participants

Participation in the MAQC main study is voluntary, and each participant is expected to cover its own costs associated with participating in the MAQC project. Each participant agrees to the conditions set in this guidance document.

4.1 RNA Providers

Ambion and Stratagene are the providers of the four RNA samples (Table 1) to be analyzed in the MAQC main study. RNA samples are provided to the test sites free-of-charge. Any questions related to RNA samples should be addressed to Ambion's Mike Wilson or Stratagene's Gavin Fischer (Table 3).

Table 3. RNA providers

Provider	Contact	Phone / Fax	E-mail	Address
Ambion	Mike Wilson	512-651-0200 X6236 / 512-651-0201	mwilson@ambion.com	2130 Woodward Street, Austin, TX 78744
Stratagene	Gavin Fischer	858-535-5400 X13075 / 858-535-0071	gavin.fischer@stratagene.com	11011 North Torrey Pines Road, La Jolla, CA 92037

4.2 Platform Providers

Microarray manufacturers are providing arrays and reagents to their test sites free-of-charges (except for Novartis, which will cover the arrays and reagent costs by itself to test the Affymetrix platform). Contact information of the representatives of platform providers is listed in Table 4.

Table 4. Platform providers

No.	Provider	Contact	Phone / Fax	E-mail	Address
1	Affymetrix	Chunmei Liu	408-731-5502 / 408-481-0422	chunmei_liu@affymetrix.com	3380 Central Expressway, Santa Clara, CA 95051
2	Agilent	Jim Collins	650-485-5048 / 650-485-8670	jim_collins@agilent.com	3500 Deer Creek Road, MS 25U-7, Palo Alto, CA 94304
3	Applied Biosystems	Lu Zhang	650-638-5045 / 650-638-6343	Lu.N.Zhang@appliedbiosystems.com	850 Lincoln Centre Drive, Foster City, CA 94404
4	Combimatrix	Mark Elliot	425-493-2285	melliott@combimatrix.com	6500 Harbour Heights Parkway, Mukilteo, WA 98275
5	Eppendorf	Francoise de Longueville	+32-81725615 / +32 81725614	delongueville.f@eppendorf.be	Eppendorf Array Technologies SA, rue du Séminaire 20A, B-5000 Namur, Belgium
6	GE Healthcare	Rich Shippy	480-722-2339	Richard.Shippy@ge.com	GE Healthcare, Molecular Diagnostics, 3200 W. Germann Rd., Chandler, AZ 85248
7	Illumina	Shawn Baker	858-202-4597 / 858-202-4680	scbaker@illumina.com	9885 Towne Centre Drive, San Diego, CA 92121
8	NCI_Operon	Ernie Kawasaki	301-435-2891 / 301-402-3134	kawasake@mail.nih.gov	Microarray Facility, NCI Advanced Technology Center, 8717 Grovemont Circle, Gaithersburg, MD 20877
9	Applied Biosystems (TaqMan)	Kathy Lee	650-554-3496	Kathy.Y.Lee@appliedbiosystems.com	850 Lincoln Centre Drive, Foster City, CA 94404
10	Genospectra (QuantiGene)	Yuling Luo	510-818-2623	yluo@genospectra.com	6519 Dumbarton Circle, Fremont, CA 94555

4.3 Test Sites

Twenty-eight (28) sites (Table 5) will test the eight microarray platforms and two alternative technologies. Full contact information of the test sites is listed in Appendix 1: MAQC_Test_Sites.pdf.

Table 5. The "official" test sites

No.	Manufacturer (Site 1)	Site 2	Site 3
1	Affymetrix*	FDA/CDER	Ambion
2	Agilent	FDA/NCTR	Icoria
3	Applied Biosystems	EPA/NHEERL	Vanderbilt Univ.
4	Combimatrix	UCSF	Stanford
5	Eppendorf	MD Anderson	CSHL
6	GE Healthcare	UMass Boston	Genus Biosciences
7	Illumina	Duke University	Burnham Institute
8	NCI (Custom arrays)	FDA/NCTR	FDA/CBER
9	AB (TaqMan)	N/A	N/A
10	Genospectra (QuantiGene)	N/A	N/A

*Three additional sites (site 4: EPA; site 5: Novartis; and site 6: UCLA/Cedars-Sinai) will test the four RNA samples with the Affymetrix platform and submit the datasets to MAQC. However, these datasets will not be included in the MAQC main manuscript.

4.4 Data Analysis Sites

Ten (10) sites have requested access to the MAQC datasets to conduct independent analyses (Table 6).

Table 6. Data analysis sites*

No.	Analysis site	Contact	Phone / Fax	E-mail	Address
1	Expression Analysis	Wendell Jones Laura Reid	919-405-2248	wjones@expressionanalysis.com; lreid@expressionanalysis.com	2605 Meridian Parkway, Durham, NC 27713
2	FDA/NCTR	Leming Shi Jim Chen	870-543-7387 / 870-543-7854	leming.shi@fda.hhs.gov; james.chen@fda.hhs.gov	3900 NCTR Road, Jefferson, AR 72079
3	Harvard	Zoltan Szallasi	617-355-2179 / 617-730-0921	zszallasi@chip.org	423 Brookline Ave. #144, Boston, MA 02215
4	NIH/NCBI	Damir Herman Tao Tao	301-594-2274 / 301-480-2290	herman@ncbi.nlm.nih.gov: tao@ncbi.nlm.nih.gov	NCBI, NLM, NIH, Building 38A, Room 6S614-L, Bethesda, MD 20894
5	NIST	Walter Liggett	301-975-2851	walter.liggett@nist.gov	Statistical Engineering Division, NIST, 100 Bureau Drive, Stop 8980, Gaithersburg, MD 20899-8980
6	SAS	Wenjun Bao Russ Wolfinger	919-531-1484 / 919-677-4444	wenjun.bao@sas.com; russ.wolfinger@sas.com	Building S-4055, SAS Campus Drive, Cary, NC 27513
7	Stanford University	Hanlee Ji Jochen Kumm	650-796-7999 / 650-724-5791	hanleeji@stanford.edu; jochen@stanford.edu	Stanford Genome Technology Center, Clark Center W300, 318 Campus Drive, Stanford, CA 94305-5440
8	UIUC	Sheng Zhong	217-333-1867	szhong@uiuc.edu	Dept of Bioengineering, 3120 DCL, MC-278, 1304 W. Springfield Ave, Urbana, IL 61801
9	UMass Boston	Rick Jensen	617-287-6032 / 617-287-6050	roderick.jensen@umb.edu	Microarray Suite, Center for Environ Health, Science and Technology, Univ of Massachusetts Boston, 100 Morrissey Blvd, S-4-023, Boston, MA 02125
10	ViaLogy	Cecilie Boysen Bud Bromley	626-296-6326 / 626-296-6329	cecilie.boysen@vialogy.com; bud.bromley@vialogy.com	2400 Lincoln Ave, Altadena, CA 91001

*All platform providers (9) are required to conduct independent data analyses; test sites (28) are encouraged to conduct data analyses.

4.5 Procedure for Including an Additional Site

There are three "official" test sites for each microarray platform (Table 5). Requests for being included in the MAQC project as an "unofficial" test site or analysis site will be

reviewed during the MAQC teleconferences. If approved, the organizers (Section 7) will ensure that the new participants are adequately briefed and agree to the conditions of the MAQC guidance document.

Test sites that are part of the main study and agree to the conditions of this guidance document will not be replaced or substituted by another test site. Circumstances may arise that could warrant a legitimate exception to this policy. In this case, all other platform providers must agree to the substitution. For example, a platform provider may request the withdrawal of a dataset from its test site if the test site failed to follow the manufacturer's SOP. An extension of the deadline for its platform to be included in the MAQC project/manuscript can be considered but it must be approved by representatives from each of the other platforms.

4.6 Procedure for Excluding a Platform or Test Site

Test sites are expected to process the samples and return the datasets promptly. A target date will be set for submitting data within 5 weeks of the date RNA samples are shipped. This target is set to allow the test sites enough time to engage in an SOP-practice phase to compare results between test sites of a platform. A test site will be excluded if its dataset is not submitted to the FDA/NCTR by the deadline set at 8 weeks after the distribution of the test RNA samples.

5. Responsibilities of Participants

1. Each platform provider will provide consistent lots of arrays and kits to test sites along with a standardized protocol on the generation of the data for the MAQC main study.
2. Each platform provider will help ensure performance capabilities of the selected test sites by providing arrays and reagents, and potential training. A pre-study phase is encouraged to allow test sites to build proficiency with the SOP that may be different from the one used at the site.
3. Each test site will be expected to practice and perform the standardized protocol and kit to ensure competency and consistency with the protocol.
4. Test sites will report target preparation yields and Bioanalyzer data on the assessment of target preparations. See Appendix 3 and Appendix 4 for requirements.
5. Test sites commit to submitting data to FDA/NCTR within 5 weeks of receiving the RNA. If a site does not submit data within 8 weeks, the site may be excluded from the study.
6. Data analysis sites will report analysis results at the face-to-face meeting in Palo Alto, CA, December 1-2, 2005.
7. Participants agree to the publication and public deposition of the submitted datasets at the time of acceptance of the initial MAQC manuscript. (For exceptions, see the data distribution section).
8. Participants agree to the confidentiality terms before accessing the MAQC datasets.

6. Confidentiality Terms for Accessing the MAQC Datasets

1. No organization should disseminate the MAQC datasets to others without written consent from the FDA/NCTR (Leming.Shi@fda.hhs.gov).
2. Prior to acceptance for publication of the main manuscript, public presentation and/or publication of the MAQC results without the consent of the MAQC are prohibited.
3. Probe sequences from Applied Biosystems and Eppendorf should only be shared within the Sequence Mapping group, whose members signed CDA's with providers.

7. MAQC Organizing Committee

The MAQC Organizing Committee (Table 7) is proposed to resolve any remaining issues not addressed in this guidance document or when it becomes unfeasible to reach consensus among all members of the MAQC e-mail distribution list (~150). The proposed Committee consists of representatives from multiple FDA Centers, EPA, non-regulatory governmental agencies, academia, RNA and platform providers.

Table 7. Members of the MAQC Organizing Committee*

No.	Name	E-mail	Organization	Function
1	Felix Frueh	felix.frueh@fda.hhs.gov	FDA/CDER	Reg gov agency
2	Jim Fuscoe	james.fuscoe@fda.hhs.gov	FDA/NCTR	Reg gov agency
3	Federico Goodsaid	federico.goodsaid@fda.hhs.gov	FDA/CDER	Reg gov agency
4	Heather Harbottle	heather.harbottle@fda.hhs.gov	FDA/CVM	Reg gov agency
5	Scott Jackson	scott.jackson@fda.hhs.gov	FDA/CFSAN	Reg gov agency
6	Scott Pine	patrick.pine@fda.hhs.gov	FDA/CDER	Reg gov agency
7	Raj Puri	raj.puri@fda.hhs.gov	FDA/CBER	Reg gov agency
8	Uwe Scherf	uwe.scherf@fda.hhs.gov	FDA/CDRH	Reg gov agency
9	Leming Shi	leming.shi@fda.hhs.gov	FDA/NCTR	Reg gov agency
10	Weida Tong	weida.tong@fda.hhs.gov	FDA/NCTR	Reg gov agency
11	David Dix	dix.david@epa.gov	EPA	Reg gov agency
12	Kathryn Gallagher	gallagher.kathryn@epamail.epa.gov	EPA	Reg gov agency
13	Damir Herman	herman@ncbi.nlm.nih.gov	NCBI	Non-reg gov agency
14	Ernie Kawasaki	kawasake@mail.nih.gov	NCI	Non-reg gov agency
15	Walter Liggett	walter.liggett@nist.gov	NIST	Non-reg gov agency
16	Zoltan Szallasi	zszallasi@chip.org	Harvard	Academia
17	Charles Wang	charles.wang@cshs.org	UCLA	Academia
18	Rick Jensen	roderick.jensen@umb.edu	UMass	Academia
19	Mike Wilson	mwilson@ambion.com	Ambion	RNA provider
20	Gavin Fischer	gavin.fischer@stratagene.com	Stratagene	RNA provider
21	Janet Warrington	janet_warrington@affymetrix.com	Affymetrix	Platform provider
22	Jim Collins	jim_collins@agilent.com	Agilent	Platform provider
23	Lu Zhang	lu.n.zhang@appliedbiosystems.com	Applied Biosystems	Platform provider
24	Andy McShea	amcshea@combimatrix.com	Combimatrix	Platform provider
25	Francoise de Longueville	delongueville.f@eppendorf.be	Eppendorf	Platform provider
26	Rich Shippy	richard.shippy@ge.com	GE Healthcare	Platform provider
27	Shawn Baker	scbaker@illumina.com	Illumina	Platform provider
28	Yuling Luo	yluo@genospectra.com	Genospectra	Platform provider

*Proposed for discussion purpose only; not confirmed by the proposed representatives.

8. Main Study Procedures

Shipment of RNA can mark the official start of the experimental processing. Therefore, RNA will not be shipped to test sites until the guidance document is agreed upon and the launch requirements for the study have been met. This should help ensure that all results will achieve the intended consistency within the MAQC guidelines. The complexity of this study justifies a written procedure for all the test sites and analysis groups to follow beyond what is provided by the individual array manufacturers.

8.1 RNA Materials and Handling

- Reference RNA product sheets will be provided by Ambion and Stratagene.
- Process for creating 25%:75% defined mixtures of the two reference RNAs is documented in Appendix 2: MAQC_RNA_Preparation_SOP.pdf.
- Ambion will provide shipment of RNA sets of 4 tubes – packaging, dry ice.

8.2 Target Preparation and Assessment

Appendix 4 (MAQC_Sample_Processing_Overview_SOP.pdf) provides procedures that apply to every test site. Included are the following elements:

- Procedures for receiving and storage of RNA references at test sites
- Quality assessment of RNA reference samples
- Target preparations and replication guidelines
- Standardized nomenclature for referencing samples
- Template for reporting quality assessment data (Appendix 3: MAQC_RNA_Quality_Report_Template.xls, sheets 2 and 3).

8.3 Convention for Naming Files

Files submitted to the MAQC project will use a defined naming convention that will code for platform, site, RNA sample, replicate, and file type (Appendix 3: MAQC_RNA_Quality_Report_Template.xls (sheet 1):

- Each platform is assigned a 3-letter code (all caps) as assigned in Table 2.
- Underscores will be used for separators after the platform code and after the site code (do not use dashes or spaces).
- Each site is assigned a number 1, 2 or 3 (Table 5).
- Each RNA is assigned an “A”, “B”, “C”, or “D” according to Table 1.
- Each replicate is identified with a 1, 2, 3, 4 or 5.
- Each platform will specify the file types to be submitted and the suffix that each receives.
- Platforms that are locked into a different file naming convention will submit a reference table to allow conversion of all data to this naming format.
- Example: AFX_2_B1.cel represents the file from platform (AFX) processed by FDA/CDER (site 2) for the first replicate of Ambion Human Brain Reference RNA (B1)

8.4 Platform-specific Standard Operating Procedures (SOP)

Array manufacturers will submit SOP's and appropriate references to the MAQC organizers prior to the study. These SOPs and references should be available for review by the MAQC participants and a period should be allowed for comments. The manufacturer's manual is not likely to be sufficient for the SOP required for the purposes of this study. For example, manuals often specify ranges and options that are available under a variety of circumstances that customers may face. Operational definitions need to be included for the specific conditions of this study. The SOP should specify which sections of the manual are to be followed.

SOP Required Elements

- SOP is submitted to MAQC for each platform.
- Quality control criteria defined for each platform that should be submitted by each test site. Cutoffs for sample quality should be specified.
- Define how deviations from the SOP will be handled and documented.
- Manufacturer-recommended procedure for data normalization and analysis (see data analysis section).

8.5 Data Submission Procedures

1. Each test site should submit its dataset to the FDA/NCTR no later than ***5 weeks after the date RNA is received***. Platform providers and test sites should work as closely as possible to avoid unnecessary delays.
2. The types of data files that will be submitted by a platform are listed in Appendix 3: MAQC_RNA_Quality_Report_Template.xls (sheet 1).
3. Quality data for RNA samples and targets should be submitted (Appendix 2).
4. Provide descriptions of any data normalizations/transformations and background subtraction.
5. Provide initial quality assessment of the submitted data.
6. Data submitted to the MAQC project will be submitted to a public repository at the time of manuscript submission, tentatively scheduled for February 2006.
7. A platform provider may request the withdrawal of a dataset from its test site if the test site failed to follow the manufacturer's SOP. However, the platform providers must substitute the withdrawn dataset with another dataset before the deadline for its platform to be included in the MAQC project/manuscript (see section on excluding and including test sites.)
8. Data should be submitted in DVD's to:

   Leming Shi
   FDA/NCTR
   Bldg 5C/Rm 109L
   3900 NCTR Road
   Jefferson, AR 72079
   870-543-7387
   leming.shi@fda.hhs.gov

8.6 Data Distribution Procedures

1. FDA/NCTR will load all datasets into its ArrayTrack software upon receipt.
2. The entire data sets from each test site will be available from ArrayTrack. The data file types are different for each platform and the file types to be submitted by each platform are defined in Appendix 3 (MAQC_RNA_Quality_Report_Template.xls, sheet 1). Each platform will specify the types of data that will be distributed by DVD to all the analysis sites. If needed, an analysis site may request other files types (e.g., image files) from the FDA/NCTR.
3. Each platform provider, RNA provider, data analysis site, and test site will be granted full access to the datasets generated in the MAQC main study after all datasets are submitted. Instructions for distributing and accessing the datasets will be provided separately by the FDA/NCTR.
4. There is no restriction on the types of analyses that will be performed by each site.
5. Each site accessing the MAQC datasets must strictly fulfill its obligations for confidentiality as set forth in this guidance document.
6. Data will be distributed, and reviewed in three phases
    1) As each dataset is submitted, it will be reviewed by Hong Fang and Leming Shi for completeness and errors. No participants will be given access to any data from any other site during this phase and no sites should share their data independently with any other sites.
    2) When all three datasets from a particular platform are submitted, the data for that platform will be released to the three test sites, so they can start performing the intra-platform comparability analyses.
    3) When datasets from all test sites have been submitted, the data will be assembled and distributed by Leming Shi on DVD to all the test sites and the analysis sites.
7. Platform providers or individual test sites have the option to withdraw from the study and retract their datasets after the second phase of the data distribution plan above but before the entire datasets are distributed to all sites (phase 3).

8.7 Data Analysis Procedures

Platform-specific data treatment

Each platform manufacturer will provide a description of its preference on how its data should be handled for consistent treatment regarding

1. Background subtraction
2. Transformation
3. Normalization
4. Data filtering
5. QC cutoffs for flags
6. QC cutoffs for failed samples
7. Platform-specific issues

Data analysis approaches

A separate document on the strategies of data analysis is being developed and will be distributed and discussed within the MAQC group while RNA samples are being processed.

8.8 Face-to-face Meeting, December 1-2, 2005, Palo Alto, CA

Results on the analysis of the MAQC datasets will be discussed.
Outlines for manuscript drafting will be discussed.
Next steps of the MAQC project will be determined.

8.9 Planning for Publication

Principles for manuscripts derived from this study

The decisions about the content of the manuscripts that derive from this study are of significant importance. It is the intent of the MAQC organizers to produce a single manuscript for rapid publication that is synchronized with the release of the datasets into the public repositories. After that date, it is expected that additional manuscripts will be generated that will be based entirely or in part on these datasets. These manuscripts may be produced by the MAQC project members or by anyone in the public.

Main manuscript

The main manuscript describes the MAQC project; QC metrics and thresholds of each platform; correlation of step-by-step QC data with overall quality of microarray data; cross-platform comparability; concordance with TaqMan and QuantiGene; relative accuracy based on titration datasets.
Drafting team: Leming Shi, Mike Wilson, Jim Fuscoe, one representative from each platform provider.
Authorship: People who contribute significantly to the MAQC project, including representatives from platform providers, RNA providers, test sites, data analysis sites. The final decision on authorship will be made by the MAQC Organizing Committee (Table 7).

Additional manuscripts

Sequence-based cross-platform mapping: It will be prepared independently of the MAQC study results. This manuscript will have independent content and can be developed and released as soon as it is finalized. Damir Herman, Zoltan Szallasi and Scott Pine will coordinate the manuscript drafting effort.
Titration Pilot: The Titration Pilot team will work on it with the titration pilot datasets.
Comparison of normalization and gene selection methods: Leming Shi and Jim Chen (FDA/NCTR) are conducting a comprehensive study on this topic and intend to use the MAQC datasets as an example.
Other proposals for manuscripts are welcome.
The authorship of these additional manuscripts will be determined by the respective manuscript drafting team.

9. Checklist of Requirements before Main Study Sample Processing Will Start

- A revised version of this guidance document will be reviewed and made effective by consensus of the MAQC group.

- Participants will agree to the scope of this project and accept the responsibilities of participants outlined within the guidance document.
- Platform providers will provide SOP's for review.

10. Timeline Target Dates

1. **August 22, 2005**: Shipment of RNA samples to test sites.
2. **September 26, 2005**: Submission of datasets to FDA/NCTR.
3. **October 3, 2005**: Distribution of datasets to analysis sites, test sites and providers.
4. **October 12, 2005**: Absolute deadline for data submission (sites excluded if data is not submitted). Submission of TaqMan data may be delayed until October 31, 2005.
5. **December 1-2, 2005**: Face-to-face meeting in Palo Alto, CA to discuss results on the analysis of the MAQC datasets; drafting of manuscripts.
6. **February 11, 2006**: Submission of manuscript(s); Submission of datasets to public repositories.
7. **October 2006**: MAQC public meeting on microarray quality control and data analysis at FDA/NCTR or FDA/CDER.
8. **December 2007**: Guidance on microarray quality control and data analysis.

11. References

1. Marshall E. Getting the noise out of gene arrays. *Science*, 306(5696), 630-631 (2004).
2. Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*, 31(19), 5676-5684 (2003).
3. Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. Independence and reproducibility across microarray platforms. *Nat Methods*, 2(5), 337-344 (2005).
4. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W. Multiple-laboratory comparison of microarray platforms. *Nat Methods*, 2(5), 345-350 (2005).
5. Bammler T and Members of the Toxicogenomics Research Consortium. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods*, 2(5), 351-356 (2005).
6. Shippy R, Sendera TJ, Lockner R, Palaniappan C, Kaysser-Kranich T, Watts G, Alsobrook J. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics*, 5(1), 61 (2004).
7. Petersen D, Chandramouli GV, Geoghegan J, Hilburn J, Paarlberg J, Kim CH, Munroe D, Gangi L, Han J, Puri R, Staudt L, Weinstein J, Barrett JC, Green J, Kawasaki ES. Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC Genomics*, 6(1), 63 (2005).
8. Carter SL, Eklund AC, Mecham BH, Kohane IS, Szallasi Z. Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics*, 6(1), 107 (2005).
9. Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z. Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res*, 32(9), e74 (2004).
10. Chen JJ, Delongchamp R, Tsai CA, Hsueh HM, Sistare F, Thompson KL, Desai VG, Fuscoe JC. Analysis of variance components in gene expression data. *Bioinformatics*, 20(9), 1436-1446 (2004).

11. Frueh FW, Huang SM, Lesko LJ. Regulatory acceptance of toxicogenomics data. *Environ Health Perspect*, 112(12), A663-A664 (2004).
12. Frueh FW, Goodsaid F, Rudman A, Huang S-M, Lesko LJ. The need for education in pharmacogenomics: a regulatory perspective. *Pharmacogenomics Journal* 5(4), 218-220 (2005).
13. Lesko LJ, Woodcock J. Translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. *Nature Reviews Drug Discovery*, 3(9), 763-769 (2004).
14. Salerno RA, Lesko LJ. Pharmacogenomic data: FDA voluntary and required submission guidance. *Pharmacogenomics*, 5(5):503-505 (2004).
15. Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ, and Casciano DA. Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics*, 6(Suppl 2), S12 (2005).
16. Shi L, Tong W, SU Z, Han T, Han J, Puri RK, Fang H, Frueh FW, Goodsaid FM, Guo L, Branham WS, Chen JJ, Xu ZA, Harris SC, Hong H, Xie Q, Perkins RG, Fuscoe JC. Microarray scanner calibration curves: characteristics and implications. *BMC Bioinformatics*, 6(Suppl 2), S11 (2005).
17. Shi L, Tong W, Fang H, Goodsaid F, Frueh FW, Han T, Fuscoe JC, and Casciano DA. QA/QC: Challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Review of Molecular Diagnostics*, 4(6), 761-777 (2004).
18. The MAQC Project: Establishing QC Metrics and Thresholds for Microarray Quality Control, Abstract accepted for oral presentation at the MGED 8 Meeting, September 11-13, 2005, Bergen, Norway. http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/docs/MAQC_MGED8Abs.pdf

12. Web Sites

1. The MAQC (Microarray Quality Control) Project: http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm
2. ArrayTrack: http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/
3. Genomics at FDA: http://www.fda.gov/cder/genomics/
4. ERCC: http://www.cstl.nist.gov/biotech/ERCC/testplan.htm
5. NIST Metrology for Gene Expression: http://www.cstl.nist.gov/div831/StucturalBiology_CARB/Genomics_Metrology.htm

13. Appendices

Accessible at: http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm
Appendix 1: MAQC_Test_Sites.pdf
Appendix 2: MAQC_RNA_Preparation_SOP.pdf
Appendix 3: MAQC_RNA_Quality_Report_Template.xls
Appendix 4: MAQC_Sample_Processing _Overview_SOP.pdf
Appendix 5: MAQC_Guidance_Data_Analysis.doc (in preparation)
Appendix 6: MAQC_SOP_Affymetrix.pdf
Appendix 7: MAQC_SOP_Agilent.pdf
Appendix 8: MAQC_SOP_AppliedBiosystems.pdf (microarray and TaqMan)
Appendix 9: MAQC_SOP_Combimatrix.pdf
Appendix 10: MAQC_SOP_Eppendorf.pdf
Appendix 11: MAQC_SOP_GEHealthcare.pdf
Appendix 12: MAQC_SOP_Illumina.pdf
Appendix 13: MAQC_SOP_NCI_Operon.pdf
Appendix 14: MAQC_SOP_Genospectra.pdf